НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.89

До захисту допущено
В. о. завідувача кафедри ММСА

О.Л.Тимощук

«___» _____ 2019 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 122 Комп'ютерні науки
на тему: «Система біометричної верифікації користувача на основі методів машинного навчання»

Виконав:
студент ІІ курсу, групи КА-83 мп
Гаврилович Марія Павлівна _____

Керівник: професор кафедри ММСА,
д.т.н., проф. Данилов В.Я.

_____

Рецензент: професор кафедри
ІБ ФТІ НТУУ «КПІ ім. І. Сікорського»,
д.т.н., проф. А.Б. Качинський

_____

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань
Студент _____

Київ
2019

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)
Спеціальність — 122 «Комп'ютерні науки»

ЗАТВЕРДЖУЮ
В. о. завідувача кафедри ММСА
О. Л. Тимощук
«___» _____ 2019 р.

ЗАВДАННЯ
на магістерську дисертацію студентці Гаврилович Марії Павлівні

1. Тема дисертації: «Система біометричної верифікації користувача на основі методів машинного навчання», науковий керівник дисертації Данилов Валерій Якович, професор кафедри ММСА, д.т.н., проф., затверджені наказом по університету від «08» листопада 2019 № 3862-с

2. Термін подання студентом дисертації: 13 грудня 2019 р.
3. Об'єкт дослідження: Біометрична верифікація користувача
4. Предмет дослідження: методи машинного навчання, якім ожна використовувати для проведення біометричної верифікації користувача
5. Перелік завдань, які потрібно розробити:
1) дослідити сучасний стан та особливості застосування методів машинного навчання для біометричної верифікації користувача;
2) розробити систему підтримки прийняття рішень для біометричної верифікації користувача;
3) обрати моделі для реалізації поставленої задачі та створити програмний продукт;
4) пошук даних для застосування в програмі;
5) зробити порівняльний аналіз моделей машинного навчання для біометричної верифікації користувача;
6) розробити стартап-проект виведення на ринок результатів дослідження;
7) розробити концептуальні висновки за результатами наукового дослідження
6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

1). Архітектури автокодувальників (рис.);
2). Графіки площі під кривою для оцінки реущльтатів математичнлї моделі (рис.);
3). Архітектури моделей використаних в програмному продукті (рис.);
4). Приклади функціонування створеного програмного продукту (рис.);
5). Таблиці у розділі стартап-проекту

7. Орієнтовний перелік публікацій:

Публікація наукової статті у фаховому журналі «_____»: Порівняльний аналіз автокодувальників для біометричної верифікації користувача за показниками грудного акселерометра.

8. Дата видачі завдання: 05 вересня 2019 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об’єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	05.09.2019—10.09.2019
2.	Перший розділ. Огляд літературно-інформаційних джерел. Постановка проблеми	10.09.2019—30.09.2019
3.	Другий розділ. Огляд автокодувальників та їх використання в задачах виявлення аномалій	01.10.2019—10.10.2019
4.	Третій розділ. Побудова системи біометричної верифікації користувача	11.10.2019—20.10.2019
5.	Четвертий розділ. Розробка стартап проекту	21.10.2019—26.10.2019
6.	Концептуальні висновки. Перспективи розвитку отриманих рішень	27.10.2019—30.10.2019

Студент М.П.Гаврилович

Науковий керівник дисертації В.Я.Данилов

РЕФЕРАТ

Магістерська дисертація: 60 с., 10 рис., 25 табл., 20 джерел.

Здійснено порівняльний аналіз використання періодичних автокодувальників як складової запропонованої автоматизованої системи підтримки прийняття рішень для біометричної верифікації користувача поведінкового типу за показниками датчика грудного акселерометра.

Метою запропонованої системи є здійснення постійної, неявної перевірки користувача на основі акселерометра, для підвищення безпеки використання програми чи пристрою, а також поліпшення взаємодії користувачів з пристроєм.

Предметом дослідження є побудування системи біометричної верифікації з допомогою методів машинного навчання.

Об’єктом дослідження є дані з пристрою користувача а саме нагрудного акселерометра.

Автокодувальники дають змогу реалізовувати однокласову класифікацію, тобто навчання без вчителя, а також за допомогою додаткових шарів в архітектурі мережі дозволяють автоматично провести генерацію признаків та кодувати вхідний зразок даних в вектор признаків, що значним чином полегшує обробку даних та робить систему більш автоматизованою. Ціллю моделі є визначити границі позитивного класу, а саме відрізняти конкретного користувача від інших, тобто розв’язувати задачу виявлення аномалій. В ході порівняльного аналізу було порівняно три типи рекурентних автокодувальників з архітектурою рекурентних вузлів типу “довгої короткочасної пам’яті” з двома методами класичного машинного навчання (однокласові опорні векторні машини та “ізоляційний” ліс), для яких необхідна була ручна генерація признаків.

Реалізація концепції використання рекурентних автокодувальників для біометричної верифікації була імплементована та протестована на відкритому наборі даних.

Використання рекурентних автокодувальників для верифікації на основі поведінкових патернів показало надійні та високоточні результати та можливість впровадження таких алгоритмів у сучасних системах безпеки. Подальше дослідження може включати використання більш досконалих моделей, таких як ансамблі автокодувальників, а також розробка більш складних векторних метрик для оцінки результатів подібних моделей.

БІОМЕТРИЧНА ВЕРИФІКАЦІЯ, РЕКУРЕНТНІ АВТОКОДУВАЛЬНИКИ, ВИЯВЛЕННЯ АНОМАЛІЙ, ВАРІАЦІЙНИЙ АВТОКОДУВАЛЬНИК

ABSTRACT

Master thesis: 60 p., 10 fig., 25 tables., 20 sources.

A comparative analysis of the use of recurrent auto-encoders as a component of the proposed automated decision support system for biometric verification of the behavioral type user from the indicators of the chest accelerometer sensor is carried out. The purpose of the proposed system is to perform continuous, implicit user verification based on the accelerometer, to improve the security of the use of the application or device, as well as improve user interaction with the device.

The goal of research is to build a continuous-based user biometric verification system based on accelerometer data using unsupervised deep learning recurrent algorithms.

The subject of the study is to build a biometric verification system using machine learning methods.

The object of the study is data from a user's device, namely a breast accelerometer.

Auto-encoders enable one-class classification, ie non-teacher training, as well as additional layers in the network architecture to automatically generate features and encode a data input into a feature vector, which greatly facilitates data processing and makes the system more automated. The purpose of the model is to define the boundaries of the positive class, namely to distinguish a specific user from others, that is, to solve the problem of detecting anomalies. The comparative analysis compared three types of recurrent auto-encoders with recurrent units of long-short term memory type with two methods of classical machine learning (one-class support vector machines and “isolation” forest) which required manual feature generation.

Proof of concepts of using recurrent autoencoders for biometric verification was implemented and tested on the open-source dataset.

Usage of recurrent autoencoders for user behavioral-based verification has shown robust and high accuracy results and the ability of implementing such algorithms in modern security systems. Further research can include using more sophisticated models such as stacked autoencoders and also propose some complex vector metric for evaluating the results of such kind of models.

BIOMETRIC VERIFICATION, RECURRENT AUTO-ENCODER, ANOMALY DETECTION, VARIATIONAL AUTO-ENCODER

ЗМІСТ

ЗМІСТ 4

ВСТУП 6

1 ОГЛЯД ПІДХОДІВ ТА АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ. ПОСТАНОВКА ЗАДАЧІ 10

1.2 Аналіз існуючих підходів та рішень 12

1.2.1 Статистичні та модельні тести 12

1.2.2 Методи машинного навчання 14

Висновки до розділу 16

2 ОГЛЯД АВТОКОДУВАЛЬНИКІВ ТА ЇХ ВИКОРИСТАННЯ В ЗАДАЧАХ ВИЯВЛЕННЯ АНОМАЛІЙ 17

2.1. Швидкий огляд внутрішнього устрою та архітектури автокодувальнику 17

Висновки до розділу 21

3 ПОБУДОВА СИСТЕМИ БІОМЕТРИЧНОЇ ВЕРИФІКАЦІЇ КОРИСТУВАЧА 22

3.1 Дані 22

3.2 Побудова моделей 23

3.3 Біометрична система персоналізації 26

3.4. Результати порівняльного аналізу з використанням рекурентних автокодувальників для побудови системи верифікації користувача 27

Висновки до розділу 30

4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ 32

4.1 Опис ідеї проекту 32

4.2 Технологічний аудит ідеї проекту 34

4.3 Аналіз ринкових можливостей запуску стартап-проекту 35

4.4 Розроблення ринкової стратегії проекту 43

4.5 Розроблення маркетингової програми стартап-проекту 45

Висновки до розділу 49

ВИСНОВКИ 50

ПЕРЕЛІК ПОСИЛАНЬ 51

ВСТУП

В наші дні телефони та велике розмаїття електронних пристроїв стають великою частиною нашого життя. Ми маємо багато інформації в наших телефонах, навіть таку важливу інформацію, як кількість кредитних карток, облікові дані деяких платіжних служб тощо. Також сьогодні дуже важливо проводити моніторинг та аналіз даних на своєму пристрої для подальшої розробки різних рішень, базованих на даних, що підвищить рівень задоволеності користувачів від використання програми та дасть змогу побудувати розумну систему прийняття рішень на пристрої. Тому дуже важливо забезпечити точну ідентифікацію користувача.

Проблема перевірки користувача в тому, що неможливо створити контрольоване завдання, оскільки ми не можемо порівняти одного конкретного користувача з усіма іншими користувачами. Тож це стає неможливою задачею, а мета - виявити межі одного хорошого класу, і всі зразки, які не належать до цього класу, будуть розглядатися як аномалії, і в контексті перевірки користувача це будуть інші користувачі. Дані з акселерометра є у форматі часових рядів, тому метою є вирішення проблеми виявлення аномалії на даних часових рядів, що є завданням більшої складності.

Посилаючись на [1] у нас є три типи системи верифікації користувача: на основі знань (пароль), на основі володіння (токен, смарт-карта) та на основі біометричних даних, які поділяються на фізіологічні (на основі фази, на основі сітківки, відбитків пальців) та поведінкових (розпізнавання ходи, динаміка натискання клавіш, рухи миші, розпізнавання підписів). Розпізнавання ходи означає знайти схему того, як людина ходить, стоїть, працює за комп’ютером та

виявляється за допомогою акселерометра. У своїй роботі ми досліджуємо систему верифікації на основі поведінки, яка визначатиме поведінкові моделі у різних видах діяльності користувачів (ходьба, стояння, робота за комп'ютером) за даними, зібраними за допомогою грудного акселерометра. Перевага підходу, заснованого на автокодувальниках, в тому, що достатньо мати дані лише з одного позитивного та негативного класу, і ми вирішуємо задачу однокласової класифікації. У нашій роботі ми пропонуємо використовувати автокодувальники різного типу для біометричної верифікації.

Актуальність роботи полягає в тому, що на сьогоднішній день різноманітні електронні пристрої відіграють значну роль в нашому житті: в них зберігається інформація про фінансові операції, власне дані про наші банківські рахунки, картки, а також велика кількість персональної інформації, заволодівши якою порушник може деанонімізувати нашу особистість та використовувати це і злочинних цілях. Тому актуальною є побудова та створення надійної системи верифікації користувача, яка буде здатна захистити дані користувача та його пристрій від зловмисників.

Метою роботи є система біометричної верифікації користувача на основі методів машинного навчання. Задачами дослідження є: дослідити можливість методами машинного навчання вирішити проблему біометричної верифікації користувача, порівняти різні підходи машинного навчання, а саме методи класичного машинного навчання та нейронні мережі та побудувати систему біометричний верифікації користувача.

Предметом дослідження є побудування системи біометричної верифікації з допомогою методів машинного навчання.

Об'єктом дослідження є дані з пристрою користувача а саме нагрудного акселерометра.

Як методи дослідження використано аналіз літературних джерел, моделювання, порівняльний аналіз та системний підхід.

У [2] автори досліджень пропонують удосконалену систему безперервної аутентифікації на основі неявної верифікації на смартфонах на основі шаблонів руху за допомогою автокодувальнику (з 1,3 та 5 шарами). Вони використовують прості автокодувальники, але звертають увагу на розробку розподіленої хмарної архітектури, щоб зробити можливими дорогі обчислення для смартфонів у хмарі, а тому зробити її швидшою та обчислювально ефективною.

Можна використовувати автокодувальники не для верифікації, а для зворотної проблеми - анонімізації для запобігання деанонімізації при надсиланні даних на недовірені ресурси [3]. У роботі [3] пропонується конкретна функція анонімізації, яка ґрунтується на додаванні регуляризації до функції втрат. Підхід глибокого навчання до перевірки користувача за допомогою RNN Deep Clockwork, запропонований у [4]. Рекурентний підхід важливий при роботі з послідовними даними та часовими рядами, такими як дані датчиків акселерометра, оскільки попередні сигнали впливають на майбутнє, але в [4] автори не вирішували задачу однокласової класифікації. Для біометричної ідентифікації можна використовувати не тільки дані акселерометра, але й дані ЕКГ. У [5] автори також використовують глибокі автокодувальники для ідентифікації особи, але з даними ЕКГ. Цей підхід до автоматичного кодування може бути корисним не тільки для персоналізації користувача, але і для пошуку аномалій у показниках здоров’я користувачів, і це вкрай важливо, оскільки за допомогою даних історії ви можете провести моніторинг та зробити кілька більш складних висновків про здоров'я користувачів в цілому, і це може допомогти у критичній ситуації.

Ідея неявної автентифікації, заснованої на поведінкових моделях, таких як поєднання всіх джерел даних, які ви можете запропонувати, як смартфон, як

використання програми, схеми дзвінків тощо, запропонована в [6]. Це корисно для смартфонів, але, наприклад, якщо у вас є такий фітнес-пристрій, як браслети, ви не матимете доступу до таких різноманітних даних користувачів, тому важливо мати можливість забезпечити можливість системи перевірки користувачів робити лише основу автентифікації на датчикові дані.

Іншим набором алгоритмів, які можуть бути використані як метод виявлення аномалії, є штучна імунна система (наприклад, позитивний і негативний клональний відбір), а в роботі [7] застосовано підхід штучної імунної системи для персоналізації користувача на основі сенсорних моделей поведінки.

1 ОГЛЯД ПІДХОДІВ ТА АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ. ПОСТАНОВКА ЗАДАЧІ

Багатьом додаткам потрібна можливість визначити, чи належить нове спостереження до того ж розподілу, що й існуючі спостереження (воно є внутрішнім), або має розглядатися як інше (воно є викидом). Часто ця здатність використовується для очищення реальних наборів даних. Необхідно розрізняти:

Детектування новизни (Novelty detection), де дані для навчання моделі не забруднюються викидами, і ми зацікавлені у виявленні аномалій в нових спостереженнях. Та детектування викидів (Outlier detection), де дані навчання містять викиди, і нам потрібно відповідати центральному режиму систему адаптації, ігноруючи спостереження з відхиленням.

Викиди є наслідком:

- помилок в даних (неточності вимірювання, округлення, неправильного запису і т.п.);
- наявності шумових об'єктів (невірно класифікованих об'єктів);
- присутності об'єктів «інших» вибірок (наприклад, показаннями зламаного датчика).

Новизна, як правило, з'являється в результаті принципово нової поведінки об'єкта. Скажімо, якщо наші об'єкти - опис роботи системи, то після проникнення в неї вірусу об'єкти стають «новизною». Ще приклад - опис роботи двигуна після поломки. Тут важливо розуміти, що «новизна» називається новизною з тієї причини, що такі описи для нас абсолютно нові, а нові вони тому, що ми не можемо в навчальній вибірці мати інформацію про всілякі зараження вірусами або всілякі поломки. Формування такої навчальної вибірки трудозатратно і часто не

має сенсу. Зате можна набрати достатньо велику вибірку прикладів нормальної (штатної) роботи системи або механізму. [8]

Розглянемо набір даних n спостережень з того самого розподілу, який описується функціями p. Розглянемо тепер, що ми додали ще одне спостереження до цього набору даних. Чи нове спостереження настільки відрізняється від інших, що ми можемо сумніватися, що воно регулярне? (тобто вона походить з того самого розподілу?) Або навпаки, чи так схоже на інше, що ми не можемо відрізнити його від оригінальних спостережень? Це питання, яке розглядається інструментами та методами виявлення аномалій.

Взагалі, мова йде про вивчення грубих, близьких кордонів, делімітуючих контур початкового розподілу спостережень, побудованих у вбудовуваному p-розмірному просторі. Тоді, якщо подальші спостереження лежать в межах підпростору, обмеженого границями, вони розглядаються як члени тієї ж популяції, що і початкові спостереження. Інакше, якщо вони лежать поза межами кордону, ми можемо сказати, що вони ненормальні з певним рівнем довіри до нашої оцінки.

Варто відзначити, що можливих постановок задач тут теж багато. Наприклад, завдання Positive-Unlabeled Classification (PU learning) - це коли частина викидів позначена (клас 1), але в інших об'єктах навчання (клас 0) також можуть міститися викиди. Наприклад, нам експерт сказав, що обладнання давало збій в такі моменти часу, але він міг помітити не все збої.

Навіть коли завдання виявлення аномалій схожі на звичайні завдання класифікації, є особливості, скажімо дисбаланс класів (наприклад, поломки обладнання відносно рідко зустрічаються).

Аномалії бувають не тільки в табличних даних, вони можуть бути в графах, часових рядах і т.д. Особливо складною проблемою є знаходження аномалій в часових рядах, в випадку коротких вибірок[9].

1.2 Аналіз існуючих підходів та рішень

1.2.1 Статистичні та модельні тести

Як правило, застосовують для окремих ознак і відловлюють екстремальні значення (Extreme-Value Analysis). Для цього використовують, наприклад, Z-value або Kurtosis measure. Приклад викидів бачимо на рисунку 1.1.

$$Z_i = \frac{|x_i - \mu|}{\sigma}$$

$$Kurtosis = \frac{1}{n} \sum_{i=1}^{n} Z_i^4$$

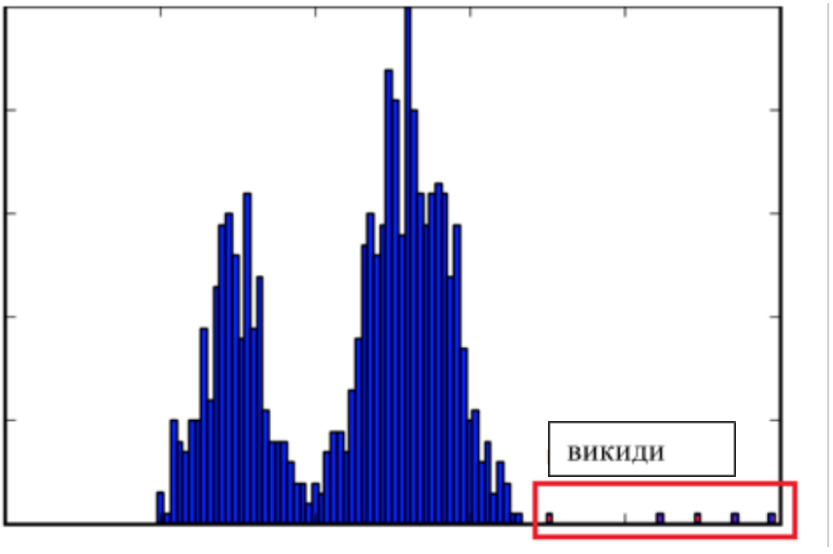


Рисунок 1.1 – Приклад викидів

Будь-який практик має який-небудь свій перевірений спосіб знаходження екстремальних значень для певних типів даних. Багато методів візуалізації, наприклад ящик з вусами, мають вбудовані засоби для детектування і показу таких екстремальних значень.

Важливо розуміти, що екстремальне значення і аномалія це різні поняття. Наприклад, в невеликій вибірці: *[1, 39, 2, 1, 101, 2, 1, 100, 1, 3, 101, 1, 3, 100, 101, 100, 100]* значення 39 можна вважати аномалією, хоча воно не є максимальним або мінімальним. Також варто відзначити, що аномалія характеризується, як правило, не тільки екстремальними значеннями окремих ознак

Ідея модельних тестів дуже проста - ми будуємо модель, яка описує дані. Точки які сильно відхиляються від моделі (на яких модель сильно помиляється) і є аномалії. При виборі моделі ми можемо врахувати природу задачі, функціонал якості і т.п.Такі методи хороші для визначення новизни, але гірше працюють при пошуку викидів. Дійсно, під час налаштування моделі ми використовуємо дані, в яких є викиди (і вона під них «заточується»).

На рис. 1.2 показано застосування модельного підходу. У нас є матриця і потрібно знайти в ній викиди. Ми використовуємо неповне сингулярне розкладення (SVD), щоб знайти матрицю невеликого рангу максимально схожу на нашу (для наочності всі числа округлені). Елементи, які сильно відрізняються від відповідних елементів матриці невеликого рангу, будемо вважати викидами.

Рисунок 1.2 – Застосування SVD для знаходження викидів в матриці

Також поширені ітераційні методи (складні в обчисленнях), метричні методи (наприклад Local Outlier Factor), методи підміни задач.

1.2.2 Методи машинного навчання

А що якщо сприйняти завдання знаходження аномалій як нове завдання машинного навчання (на відміну від класифікації і кластеризації) ?

Найпопулярніші алгоритми (є реалізація навіть в scikit-learn) такі:

- метод опорних векторів для одного класу (OneClassSVM);
- ізолюючий ліс (IsolationForest);
- еліпсоїдальної апроксимація даних (EllipticEnvelope).

Перший метод - це звичайний SVM(Support Vector Machines), який відокремлює вибірку від початку координат. Ідея трохи сумнівна, але виявилася досить працездатною. Тут правда не так багато різноманітності у виборі

параметрів, як при вирішенні задач класифікації, оскільки в якості ядра підійде лише rbf (радіально базисні функції), всі інші ядра показують просто жахливий результат. Цікаво, що багато років завдання детектування поломок складних механізмів вирішувалися саме за допомогою OneClassSVM, чомусь без розгляду альтернатив. Корисно пам'ятати, що OneClassSVM це скоріше алгоритм пошуку новизни, а не викидів, тому що «Заточується» під навчальну вибірку.

Важливі параметри реалізації sklearn.svm.OneClassSVM:

а) kernel - ядро (лінійне: linear, поліноміальний: poly, радіальні базисні функції: rbf, сигмоїдальна: sigmoid, своє заданий);

б) nu - верхня межа на% помилок і нижня на% опорних векторів (0.5 за замовчуванням);

в) degree - ступінь для поліноміального ядра;

г) gamma - коефіцієнт для функції ядра (1 / n_features за замовчуванням);

д) coef0 - параметр у функції поліноміального або сигмоїдального ядра [10].

Ізолюючий ліс (Isolation Forest) - це одна з варіацій ідеї випадкового лісу. Як завжди: проста і надійна. Складається з дерев, кожне дерево будується до вичерпання вибірки. Для побудови розгалуження в дереві: вибирається випадкова ознака і випадкове розщеплення

Для кожного об'єкта міра його нормальності - середнє арифметичне глибин листя, в які він потрапив (відокремився).

Логіка алгоритму проста: при описаному «випадковому» способі побудови дерев викиди потраплятимуть в листя на ранніх етапах (на невеликій глибині дерева), тобто викиди простіше «ізолювати» (нагадаємо, що дерево будується до тих пір, поки кожен об'єкт не виявиться в окремому аркуші). Алгоритм добре відловлює саме викиди (див. рис. 1.3).

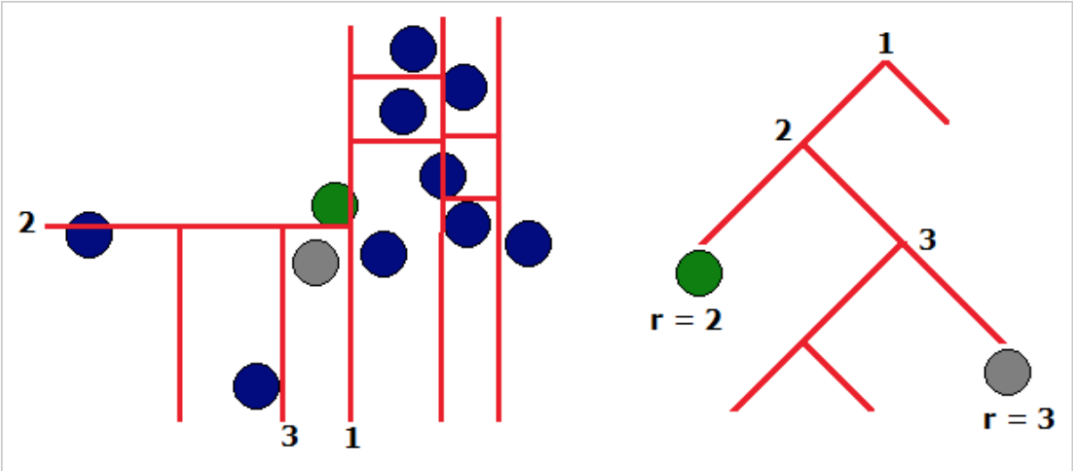


Рисунок 1.3 – Обчислення оцінки аномальності в ізолюючому лісі

Важливі параметри реалізації sklearn.ensemble.IsolationForest

- n_estimators - число дерев;
- max_samples - обсяг вибірки для побудови одного дерева (якщо дійсне число, то відсоток всієї вибірки);
- contamination - частка викидів в вибірці (для вибору порога);
- max_features - число (або%) ознак, які використовуються при побудові одного дерева (поки працює тільки зі значенням 1.0);
- bootstrap - включення режиму бутстрепа при формуванні підвибірки [11].

Еліпсоїдальна апроксимація даних - з назви зрозуміло, що група точок моделюється як внутрішність еліпсоїда. Метод добре працює тільки на одномодальних даних, а зовсім добре - на нормально розподілених. Ступінь новизни тут фактично визначається по відстані Махалонобіса.

Висновки до розділу

Проведено огляд підходів та алгоритмів машинного навчання, які застосовуються для визначення аномалій.

Розглянуто статистичні підходи, матричні підходи та моделі машинного навчання, а саме ізоляціний ліс, однокласові опорні векторні машини та еліпсоїдальну апроксимацію даних.

Ізоляційний ліс та однокласові опорні векторні машини буде використано для подальшого порівняння з методами глибинного навчання.

2 ОГЛЯД АВТОКОДУВАЛЬНИКІВ ТА ЇХ ВИКОРИСТАННЯ В ЗАДАЧАХ ВИЯВЛЕННЯ АНОМАЛІЙ

2.1. Швидкий огляд внутрішнього устрою та архітектури автокодувальнику

Архітектура автокодувальнику містить дві основні частини: енкодер і декодер.

Автокодувальники навчились створювати реконструкцію вхідного вводу. Метою є мінімізація помилки відновлення (формула 2.1):

$$E = \sum_{i=1}^{n} \| x_i - d_{\varphi}(e_{\theta}(x_i) \|, \tag{2.1}$$

де $x_1 \dots x_n$ - це рядки даних, d - декодер і e - кодер з деякими параметрами φ і θ.

Енкодер кодує вхід у деякому нижньому або більш великому просторі. Він не може бути просто скопійований з вводу на вихід, тому що ми ставимо деякі обмеження, як, наприклад, нижчий розмір внутрішніх шарів і для можливості декодера відтворити вихід, кодер повинен знайти та витягнути деякі змістовні шаблони та особливості. Мета декодера - відтворити зразок із закодованого прикладу.

Алгоритм навчання автокодувальнику - це оновлення параметрів декодера та енкодера, використовуючи алгоритми на основі градієнтного спуску для мінімізації (1) [12]. деякий поріг ε та обчислити похибку відновлення для вхідних даних, і якщо похибка відновлення для деякої точки даних перевищує поріг, ми вважаємо, що ця точка даних є аномалією.

Налаштування порогу не визначено строго і залишається для дослідника індивідуальним, виходячи з конкретних особливостей проблеми, яка потребує вирішення.

Існує кілька типів:

1. Неповний автокодувальник (коли розмірність внутрішніх шарів менше вхідних), кодер вирішує задачу щодо зменшення розмірності.

2. Роздріжений (коли розмірність внутрішніх шарів більша, ніж вхідні), що додає певний штраф до помилки відновлення.

3. Зашумовуючий (додавання деякого шуму до введення та введення його в помилку відновлення). Це мінімізує не (2.1), але $\sum_{i=1}^{n} \| x_i - d\,(e\,(\hat{x}_i)\,\| \|$, де $\hat{x}_i$ є x з деяким шумом.

4. Контрактивний (додати Штраф Ω (h) - квадратна норма Фробеніуса (сума квадратних елементів) матриці Якобіана часткових похідних, пов'язаних з функцією кодера): $\Omega\,(h) = \lambda \left|\left| \frac{\partial f\,(x)}{\partial x} \right|\right|$ [9] .

5. Варіаційні автокодувальники (variational autoencoder - VAE), які оптимізують ймовірність відновлення. VAE - це генеративна модель, яка намагається реконструювати параметри розподілу ймовірностей виходу.

Архітектуру автокодувальників зображено на рис. 2.1 нижче

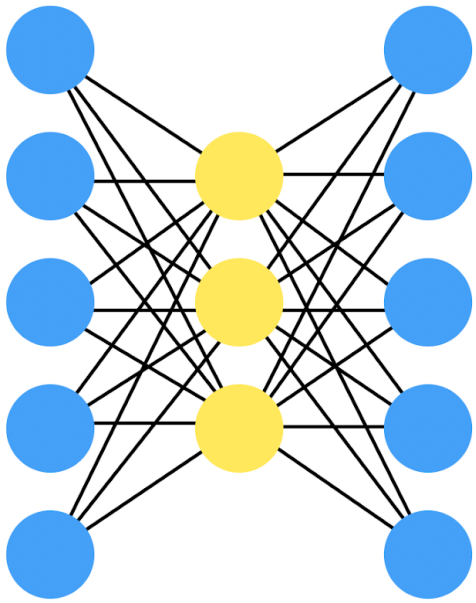


Рисунок 2.1 – Архітектура неповного автокодувальника

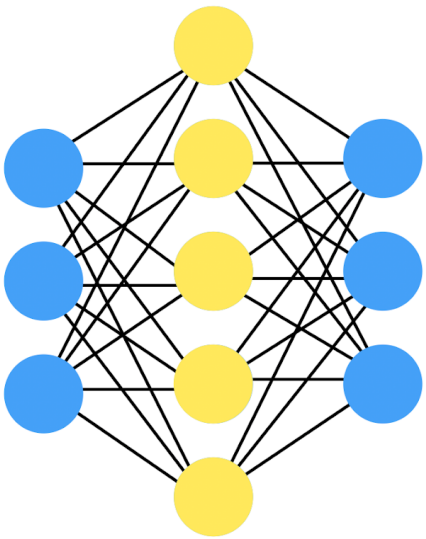


Рисунок 2.2 – Архітектура розрідженого автокодувальника

Варіаційні автокодувальники спрямовані на імовірнісну графічну модель.

Варіаційні автокодувальники не мінімізують помилку відновлення, але оптимізують ймовірність відновлення. Зовсім інший підхід, у нас все ще є кодер і

декодер, але декодер не розшифровує, але зразки прикладів даних з деякого імовірнісного розподілу з деякими параметрами і кодер переносить зразки даних у латентний імовірнісний простір, тому в основі функції втрати для варіаційної Для автокодувальнику існує дивергенція Кульбака – Лейблера, яка показує різницю між різними ймовірнісними розподілами [13]. Архітектуру варіаційного автокодувальника зображено на рис. 2.3 нижче. На ньому ми бачимо що латентний простір складається з вектора середніх значень та стандартного відхилення, тобто варіаційний автокодувальник намагається підібрати параметри для певного нормального розподілу.

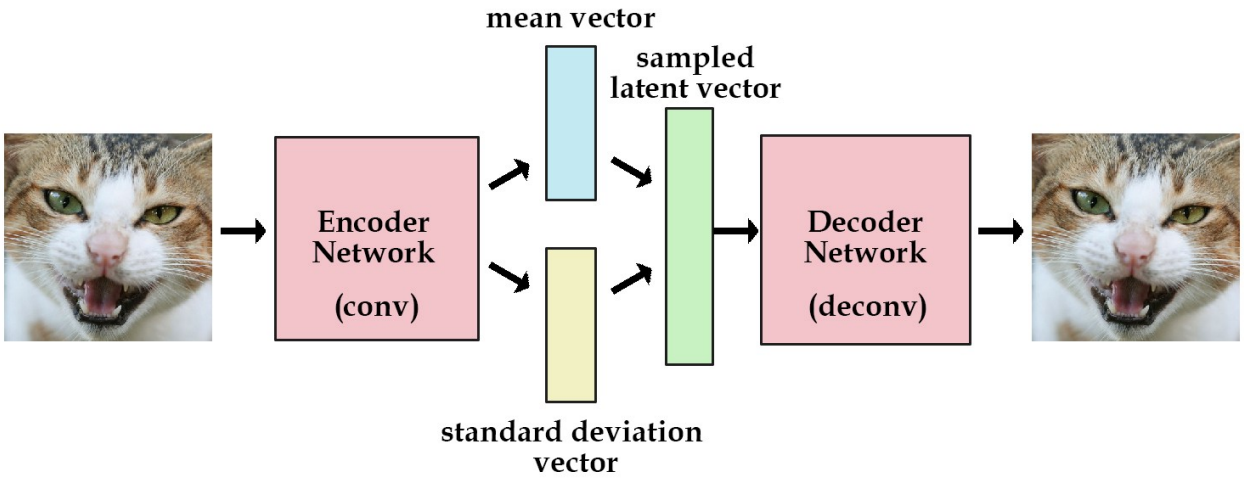


Рисунок 2.3 – Архітектура розрідженого автокодувальника

У дослідженні [10] авторів є те, що можливо комбінувати рекурентну нейронну мережу та варіаційний автокодувальник.

Отже, на основі попереднього розгляду мета полягає в тому, що для вирішення проблем персоналізації користувача на основі послідовних даних грудного акселерометра важливо використовувати періодичні архітектури для врахування попередніх значень, а через високу складність вирішення класифікаційної проблеми розрізнення користувачів найкращим підходом є

визначити межі класу "свого" (користувач у даному випадку), і для цього нам потрібні методи без машинного навчання. Отже, щоб разом використовувати рекурентність та навчання без вчителя, відповідь - це рекурентні автокодувальники.

Висновки до розділу

У даному розділі було розглянуто основні етапи роботи автокодувлаьника та основні можливі архітектури.

В подальшому дослідженні використовувалися три типи автокодувальників, а саме варіаційний, контрактивний та неповний.

3 ПОБУДОВА СИСТЕМИ БІОМЕТРИЧНОЇ ВЕРИФІКАЦІЇ КОРИСТУВАЧА

3.1 Дані

Набір даних взяли з сховища UCI - архіву загальнодоступних датасетів для цілей машинного навчання.

Набір даних містить дані акселерометра від 15 добровольців. Дані містять значення осей x, y, z [14]. Частота вибірки акселерометра становить 52 Гц, тому ми маємо 52 ряди значень щосекунди. Також дані містять 7 міток діяльності (стояти, ходити, ходити по сходах вгору і вниз, стояти та ходити, сходити вгору і вниз по сходах, ходити і розмовляти з ким-небудь, розмовляти). Як і в [15] перевірка користувача після вирішення завдання розпізнавання людської діяльності (human activity recognition - HAR) на патернах ходьби, ми також використовуємо алгоритм перевірки користувача лише за певною схемою руху. Тож насправді ми виявляємо, в який спосіб хтось ходить, ходить і розмовляє з ким-небудь і т.д.

Для моделей глибокого навчання ми розділили дані, які перекриваються на 50-відсотків на вікнах довжиною 52.

Кількість зразків для тестового та тренувального датасету, наведені у таблиці 3.1 та таблиці 3.2 нижче.

Таблиця 3.1 - Кількість зразків для тестового та тренувального датасету для 1-7 користувача

User	1	2	3	4	5	6	7
Train	866	1136	1072	812	797	1133	842
Test	428	561	529	400	393	559	416

Таблиця 3.2 - Кількість зразків для тестового та тренувального датасету для 8-15 користувача

User	8	9	10	11	12	13	14	15
Train	1133	929	1134	1394	1254	470	1361	1328
Test	559	458	559	688	619	232	671	655

3.2 Побудова моделей

Ми будуємо автокодувальники на мові програмування Python на бібліотеці Кераса та Tensorflow backend [16].

Ми порівнюємо три типи автокодувальників:

1. автокодувальник короткої довготривалої пам'яті (long short term memory - LSTM);
2. контрактивний автокодувальник LSTM;
3. неповний автокодувальник LSTM.

Як функція втрат використовувалася середня абсолютна помилка. Тренування проходило в 10 епох, 32 - розміри батчу та оптимізатор Адам.

Порогова формула:

$$threshold = \sum_{i=1}^{n} MAE_i/n + std(MAE_i)$$

де MAE це середня абсолютна помилка між зразком та тим, що передбачила модель, std - стандартне відхилення та n число зразків в тренувальній вибірці.

Архітектура автокодувальників показана на рис. 3.1.

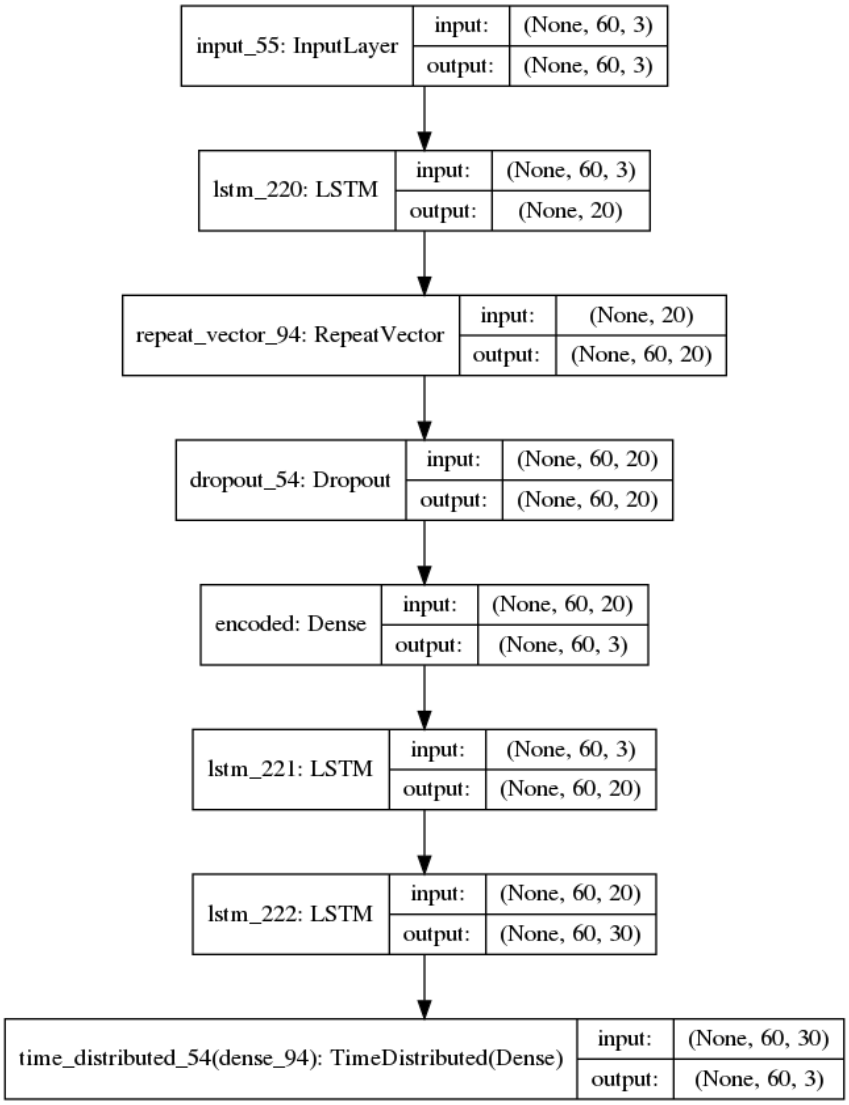
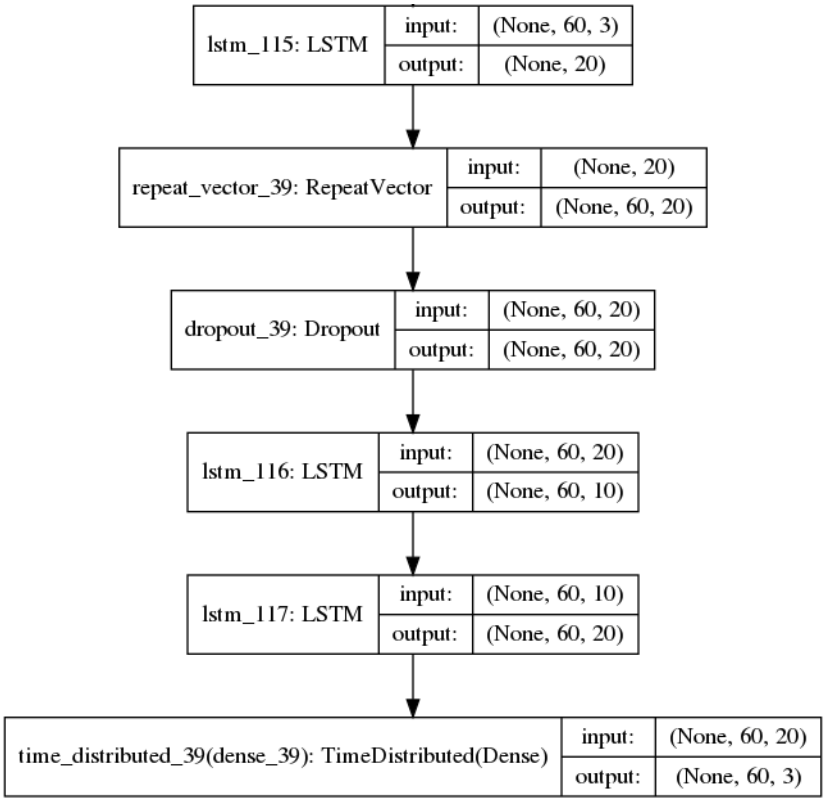
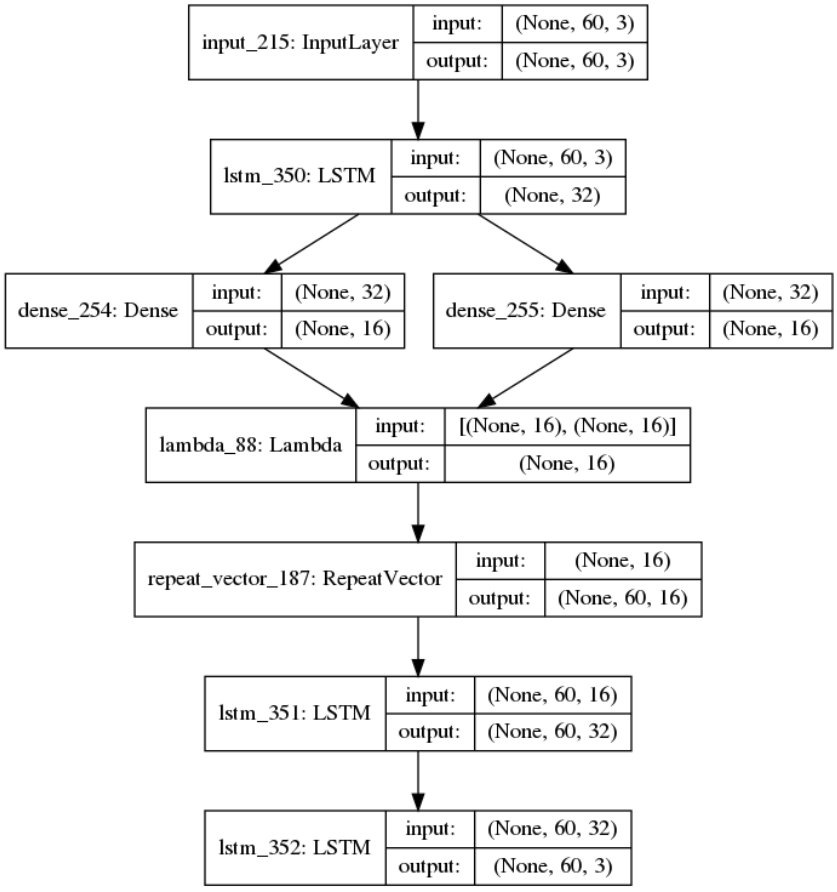


Рисунок 3.1 - Архітектура a) контрактивного автокодувальнику b) варіаційного автокодувальнику, c) неповного автокодувальнику, аркуш 1



b)

Рисунок 3.1 - Архітектура a) контрактивного автокодувальнику b) варіаційного автокодувальнику, c) неповного автокодувальнику, аркуш 2



c)

Рисунок 3.1 - Архітектура a) контрактивного автокодувальнику b) варіаційного автокодувальнику, c) неповного автокодувальнику, аркуш 3

Моделі на основі автокодувлаьників були порівняні з класичними методами машинного навчання, а саме однокласовими опорними векторними машинами та ізоляційним лісом. Для останніх двох методів використовувалися набори признаків, а не необроблені дані. Признаки (як і в часовому так і в частотному домені) були взяті як в [17].

3.3 Біометрична система персоналізації

Отже, пропонуємо систему неперервної аутентифікації (рис. 3.2).

1. Збір даних акселерометра (або інших даних датчиків, які описують закономірності руху) для різних типів діяльності.
2. Розпізнавання діяльності людини та розділення даних на частини за видами діяльності.
3. Тренування моделі та налаштування параметрів.
4. Вибір найбільш відповідної моделі для конкретного користувача на основі обраних метрик.

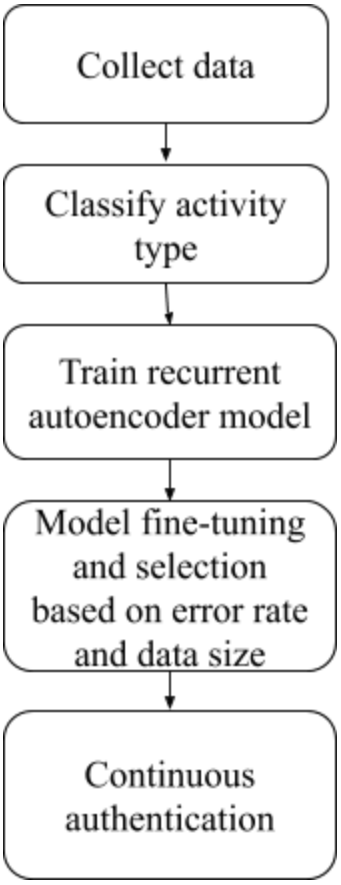


Рисунок 3.2 - Схема системи біометричної верифікації користувача

У якості метрики оцінювання моделі було обрано «recall» для позитивного («свого») класу, а також область під кривою (area under curve - AUC) для правильного порівняння моделей між собою, незалежно від порогу класифікації.

3.4. Результати порівняльного аналізу з використанням рекурентних автокодувальників для побудови системи верифікації користувача

Recall для кожного користувача та порівняння з однокласними опорними векторними машинами та ізоляційним лісом наведені в таблиці 3.3 нижче, де AE - автокодувальник, VAE - варіаційний автокодувальник, CAE - контрактивний автокодувальник, OCS - однокласовий SVM, IF - ізоляційний ліс.

Таблиця 3.3- “Recall” для позитивного класу

User/Model	LSTM AE	LSTM VAE	LSTM CAE	OCS	IF
User1	0.99	0.99	0.99	0.63	0.74
User2	0.99	0.99	0.66	0.42	0.73
User3	0.62	0.96	0.86	0.67	0.74
User4	0.44	0.58	0.66	0.27	0.74
User5	0.94	0.94	0.95	0.67	0.78
User6	0.87	0.91	0.91	0.49	0.74
User7	0.91	0.92	0.90	0.65	0.65
User8	0.97	0.97	0.97	0.39	0.83

Продовження таблиці 3.3

User9	0.50	0.55	0.56	0.27	0.76
User10	0.94	0.97	0.97	0.29	0.83
User11	0.98	0.98	0.98	0.68	0.77
User12	0.58	0.41	0.46	0.40	0.84
User13	0.77	0.27	0.59	0.20	0.88
User14	0.76	0.73	0.82	0.32	0.72
User15	0.96	0.98	0.98	0.66	0.79
Average	0.814	0.81	0.82	0.47	0.77

Також через не дуже велику різницю значення recall у моделях автокодувальника, ми порівнюємо їх і з площею під кривою. Крива ROC та значення AUC, показані на рис. 3.3 і рис.3.4.

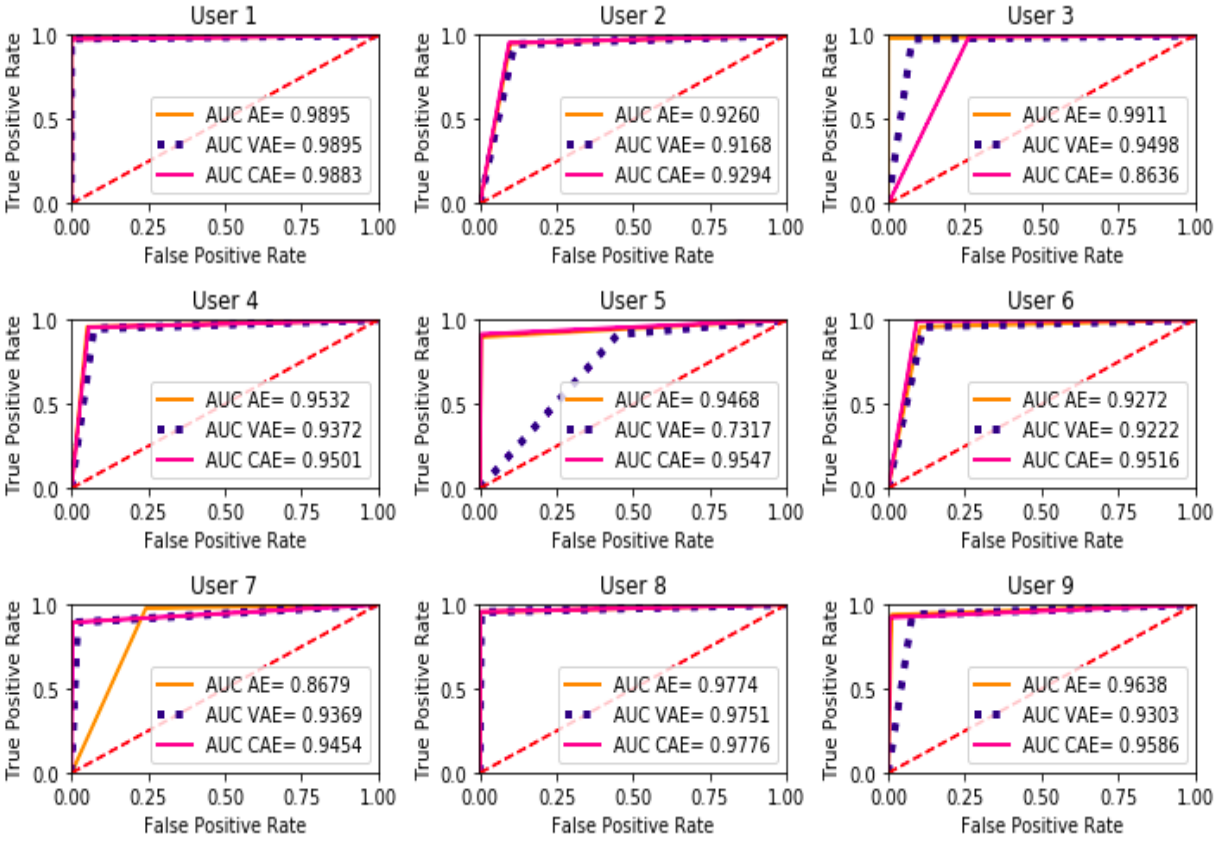


Рисунок 3.3 - ROC крива та AUC значення для моделей на базі автокодувальника (користувачі 1-9)

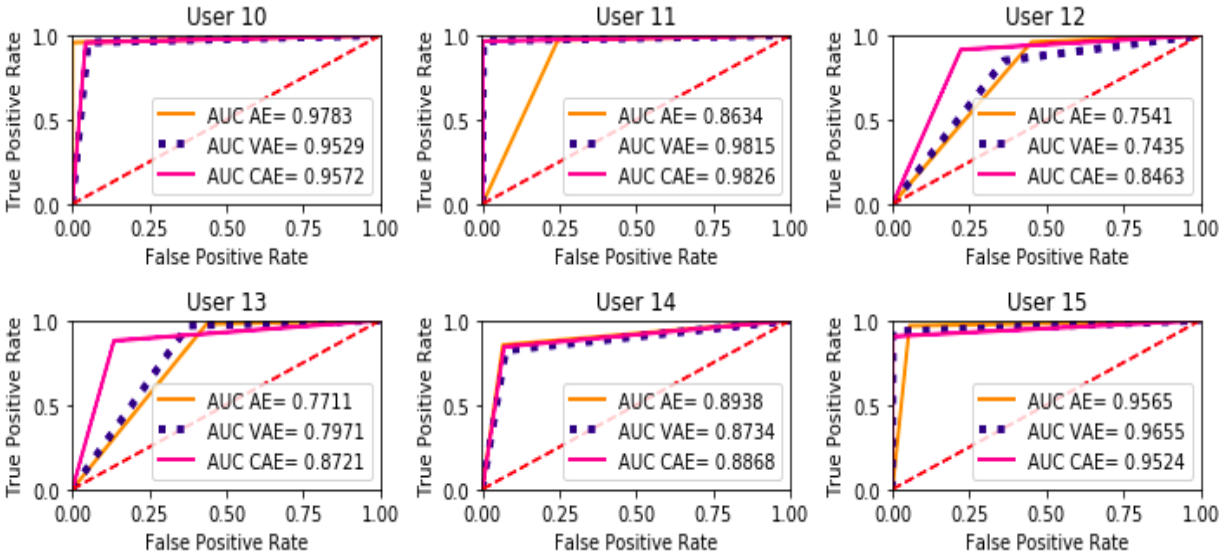


Рисунок 3.4 - ROC крива та AUC значення для моделей на базі автокодувальника (користувачі 10-15)

Висновки до розділу

В даному розділі було запропоновано систему підтримки прийняття рішень для біометричної верифікації користувача, а також реалізація модуля верифікації користувача за допомогою різноманітних методів машинного навчання.

Було проведено порівняльний аналіз методів глибинного навчання з класичними методами машинного навчання, де автокодувальники як представники групи алгоритмів глибинного навчання показали більш надійні та високі показники згідно метрик. якими оцінювалися моделі.

Порівнювалися автокодувальники трьох типів, з яких найкраще себе показали варіаційний та контрактивний. Проте варто зважати і на інші параметри,

які мають вплив на результат, такі як кількість даних по кожній активності по кожному користувачу, особисті характеристики користувача і т.п.

4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ

У даному розділі буде розглянуто ключові особливості розробленої системи як стартап-проекту. Проект розглядатиметься як система неперервної персоналізації юзера за показниками датчика акселерометра.

4.1 Опис ідеї проекту

Спочатку проаналізуємо та подамо у вигляді таблиці зміст ідеї стартап-проекту, можливі напрямки застосування та основні вигоди, які може отримати користувач товару. Ці характеристики стартап-проекту зображено в таблиці 4.1.

Таблиця 4.1 - Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Програмний додаток для персоналізації юзера по показникам датчикам акселерометра.	1. Застосування як системи неперервної автентифікації на фітнес пристроях.	Постійна неявна автентифікація.
	2. Застосування як системи неперервної автентифікації на смартфонах.	Постійна неявна автентифікація на смартфоні, що дає змогу запобігти втрати персональних даних.

Тепер зробимо аналіз потенційних техніко-економічних переваг ідеї порівняно із пропозиціями конкурентів. Результати аналізу зображено в таблиці 4.2.

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характери-тики ідеї	Товари/концепції конкурентів			W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Конкурент 1	Конкурент 2			
1.	Ціна	5000$/ рік	6000$/ рік	3000$/ рік		+	
2.	Прибутки	10000$/ рік	15000$/ рік	20000$/ рік	+		
3.	Динаміка галузі	Швидка	Швидка	Швидка		+	
4.	Постійні витрати	10000$/ рік	20000$/ рік	15000$/ рік			+
5.	Змінні витрати	5000$ -7000$/ рік	4000$ -6000$/ рік	8000$ -10000$/ рік		+	
6.	Патенти на продукти	Немає	Патент на кож-ний проект	Декілька патентів на винахід	+		
7.	Гнучкі ціни	Ціна єдина, є безкоштовна версія	Ціна варіюється з року в рік	Ціна єдина			+
8.	Законодавчі обмеження	Немає	Немає	Обмеження по зберіганню особистих даних			+

4.2 Технологічний аудит ідеї проекту

Визначимо технологічну здійсненність ідеї проекту за допомогою аналізу таких складових, як технології, за якою буде виготовлено товар згідно ідеї проекту, існування таких технологій, чи їх необхідно розробити / зробити, доступність таких технологій авторам проекту. Результати даного аналізу зображено в таблиці 4.3.

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
Програмний додаток для неперервної неявної автентифікації по показникам датчика.	Технологія проектування та розробки системи з допомогою штучних імунних систем (алгоритм позитивної та негативної клональної селекції)	Ні	Потрібно програмно реалізувати алгоритми штучних імунних систем
	Технологія проектування та розробки з допомогою Deep Learning Autoencoders.	Так	Доступні бібліотеки та фреймворки на основі яких треба буде розробити модель Autoencoders.
	Технологія проектування та розробки з допомогою ІАД	Так	Дані технології доступні.
Обрана технологія реалізації ідеї проекту: технологія Deep Learning Autoencoders.			

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Проведемо аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку. Результати даного аналізу зображено в таблиці 4.4.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	*Показники стану ринку (найменування)*	*Характеристика*
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн/ум.од	300 000
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Висока точність автентифікації
5	Специфічні вимоги до стандартизації та сертифікації	Можливість видалення особистих даних за вимогою
6	Середня норма рентабельності в галузі (або по ринку), %	150

Таким чином, за попереднім оцінюванням, ринок є привабливим для входження.

Надалі визначимо потенційні групи клієнтів, їх характеристики, та сформуємо орієнтовний перелік вимог до товару для кожної групи. Ці дані зображено в таблиці 4.5.

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	*Потреба, що формує ринок*	*Цільова аудиторія (цільові сегменти ринку)*	*Відмінності у поведінці різних потенційних цільових груп клієнтів*	*Вимоги споживачів до товару*
1	Неявна неперервна автентифікація в прошивках фітнес-пристроїв	Малий та середній бізнес	Малому бізнесу буде потрібно буде забезпечити потужності для зберігання показників датчиків та додаток для зв’язки з мобільним пристроєм.	Клієнти мають погодитися з використанням їх особистих даних
2	Неявна неперервна автентифікація в прошивках смартфонів	Малий бізнес	У малого бізнесу може виникнути потреба у наявності датчика в смартфоні та його інтеграцією з іншими системами смартфона.	Клієнти мають погодитися з використанням їх особистих даних

Після визначення потенційних груп клієнтів проведемо аналіз ринкового середовища: складемо таблиці факторів, що сприяють ринковому впровадженню проекту (таблиця 4.6), та факторів, що йому перешкоджають (таблиця 4.7).

Таблиця 4.6 – Фактори загроз

№ п/п	*Фактор*	*Зміст загрози*	*Можлива реакція компанії*
1	Відсутність попиту	Відсутня достатня кількість пристроїв на ринку, дороге впровадження.	Акцентувати увагу на клієнтах, що вже скористалися продуктом, якщо такі є, навести інфографіку результативності (очікувану), запропонувати знижку потенційному клієнту в рамках тендеру.
2	Неточне розпізнавання	Схожі поведінкові патерни в різних користувачів можуть призвести до збоїв в автентифікації.	Розробка і випуск оновлення алгоритму системи, де виправлена ця проблема.

Таблиця 4.7 – Фактори можливостей

№ п/п	*Фактор*	*Зміст можливості*	*Можлива реакція компанії*
1	Кобрендінг	Пропозиція від певної компанії, що спеціалізується на системах неявної неперервної автентифікації, розробити спільний продукт	Виділення частини штату на реалізацію проекту, підготовка акційних пропозицій по переходу на новий продукт існуючим клієнтам.

Надалі проведемо аналіз пропозиції: визначимо загальні риси конкуренції на ринку. Результати даного аналізу зображені в таблиці 4.8.

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	*В чому проявляється дана характеристика*	*Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)*
1. Чиста конкуренція	Гравці ринку не мають явних переваг один над одним	Більш вигідні умови на тендерах, агресивний маркетинг
2. Регіональна конкуренція	Гравці ринку – інтернаціональні підприємства	Вихід на ті ринки, які ще не зайняті конкурентами
3. Внутрішньогалузева конкуренція	Гравці ринку знаходяться в одній галузі – розробці ПЗ	
4. Товарно-видова конкуренція	Усі продукти гравців ринку мають одне призначення	Розробка найбільш інтуїтивного інтерфейсу, розробка унікальних мовленнєвих пакетів, оптимізація алгоритмів (щоб аналіз проходив швидше, ніж у конкурентів)
5. Конкурентні переваги нецінові	Продукти відрізняються гнучкістю, функціоналом (незначно) і надійністю.	У маркетингу неявно порівнювати власний продукт з іншими, робити вигідні цінові пропозиції
6. Марочна конкуренція	Значна увага приділяється бренду, що розробив продукт	Кобрендінг

Тепер визначимо та обґрунтуємо фактори конкурентоспроможності, які зображені в таблиці 4.9.

Таблиця 4.9 – Обґрунтування факторів конкурентоспроможності

№ п/п	*Фактор конкурентоспроможності*	*Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)*
1	Інтеграція	Продукт передбачає прошивку під мобільні операційні системи, а також інтерфейс для інтеграції з іншими пристроями.
2	Гнучкість	Кожен замовник має можливість замовити розширення функціоналу продукту під його конкретні задачі
3	Неявна неперервна автентифікація	Особисті дані користувачів, а також фінансові операції будуть в безпеці.

Таблиця 4.10 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	Динаміка галузі, продуктова лінія, бар’єри проникнення	Наявність товарних знаків, доступ до ресурсів, патенти на продукти	Концентрація постачальни-ків, диференціація витрат	Рівень чутливості до зміни цін, прибутки, контроль якості	Ціна, лояльність споживачів
Виснов-ки:	Конкуренція не є інтенсивною, адже даний ринок невеликий.	Для входу на ринок необхідно створити товарний знак та написати бета-версію програмного продукту. На даний момент потенційних конкурентів немає.	Постачальники не диктують умови роботи на ринку, бо програмному продукту не потрібно постачання.	Клієнти диктують умови роботи на ринку, бо вони є єдиним джерелом прибутку компанії.	При наявності товарів замінників необхідно буде зменшувати ціну програмного продукту чи створювати ПЗ для інших технічних систем.

За визначеними факторами конкурентоспроможності проведемо аналіз сильних та слабких сторін стартап-проекту. Результати даного аналізу зображено в таблиці 4.11.

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін системи «BioM»

№ п/п	*Фактор конкуренто-спроможності*	*Бали 1-20*	*Рейтинг товарів-конкурентів у порівнянні з BioM*						
			–3	*–2*	*–1*	*0*	*+1*	*+2*	*+3*
1	Інтеграція	15					*		
2	Гнучкість	18			*				
3	Неявна неперервна автентифікація	18			*				

Тепер проведемо SWOT-аналіз на основі виділених загроз і можливостей, та сильних і слабких сторін проекту. SWOT-матриця зображено в таблиці 4.12.

Таблиця 4.12 – SWOT-аналіз стартап-проекту

Сильні сторони: Неявна неперервна автентифікація	Слабкі сторони: Інтеграція має пройти із залученням розробників на стороні замовника
Можливості: Кобрендінг	Загрози: Неточність розпізнавання, відсутність попиту, порушення прав конфіденційності споживачів

На основі SWOT-аналізу розробимо альтернативи ринкової поведінки для виведення стартап-проекту на ринок та орієнтований оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок. Дані альтернативи зображено в таблиці 4.13.

Серед даних альтернатив було обрано другу альтернативу, адже строки її реалізації є майже найменші найменші та є ймовірність отримання ресурсів.

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

№ п/п	*Альтернатива (орієнтовний комплекс заходів) ринкової поведінки*	*Ймовірність отримання ресурсів*	*Строки реалізації*
1	Реалізація низькорівневого інтерфейсу для смарт-пристроїв	Середня	18 місяців
2	Створення неперервної неявної автентифікації на хмарних сервісах	Висока	13 місяці
3	Розробка MVP	Висока	12 місяців

4.4 Розроблення ринкової стратегії проекту

Для розроблення ринкової стратегії першим кроком необхідно описати цільові груп потенційних споживачів, які можна побачити в таблиці 4.14.

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтов-ний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1.	Малий бізнес	Середня	5-10 підприємств в рік	Слабка	Середня
2.	Середній бізнес	Готові	5-10 підприємств в рік	Слабка	Середня
3.	Великий бізнес	Готові	3-5 закладів в рік	Висока	Складна
Було обрано цільову групу підприємств групи малого бізнеса.					

Для роботи в обраних сегментах ринку необхідно сформувати базову стратегію розвитку, яку зображено в таблиці 4.15.

Таблиця 4.15 – Визначення базової стратегії розвитку

Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Концентрація на потребах одного цільового сегменту – фітнес-пристроях.	Створений продукт є інноваційним.	Стратегія спеціалізації.

Наступним кроком є вибір стратегії конкурентної поведінки, яку зображено в таблиці 4.16.

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохід-цем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
Так.	Компанія буде шукати нових споживачів, але і буде намагатися забирати існуючих у конкурентів.	Компанія буде копіювати характеристики конкурентів.	Стратегія заняття конкурентної ніші.

Тепер розробимо стратегію позиціонування, що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект. Її зображено в таблиці 4.17.

Таблиця 4.17 – Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
Розпізнавання особи за голосом має бути точним.	Проведення крупних оновлень (оптимізація розрахунків), створення додаткового функціоналу.	Товар є є інноваційним (в тренді).	Швидкий та зручний доступ до приватного контенту, додаткова система аутентифікації, програма працює в режимі онлайн.

4.5 Розроблення маркетингової програми стартап-проекту

Сформуємо маркетингову концепцію товару, який отримує споживач. В таблиці 4.18 зображено результати попереднього аналізу конкурентоспроможності товару.

Таблиця 4.18 – Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1.	Точне розпізнавання особи.	Неявна неперервна автентифікація по паттернам.	Доступність для компаній з невеликим капіталом.

Надалі розробимо трирівневу маркетингову модель товару: уточнимо ідею продукту, його фізичні складові, особливості процесу його надання. Дана модель зображена в таблиці 4.19.

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові
I. Товар за задумом	Програмний продукт – система неявної неперервної автентифікація, що дозволяє не допускати створення деяких операцій з особистими та фінансовими даними при підозрілих патернах.
II. Товар у реальному виконанні	Властивості / характеристики: 1. Можливість збору даних датчика та аналіз патернів. 2. Постійна неявна автентифікація, що дозволяє завжди слідкувати чи не відбувається ніяких зловмисницьких дій. 3. Можливість пройти аутентифікацію іншим, альтернативним, способом
	Якість: програмний продукт пройшов всі етапі тестування та готовий до використання.
	Публічне API на хмарному сервісі.
	Марка: назва організації-розробника «MG», назва товару «ConAA».
III. Товар із підкріпленням	Відділ розробки підтримує життєдіяльність ПЗ.
Захист програмного продукту буде організовано за допомогою ноу-хау.	

Тепер визначимо цінові межі, якими необхідно керуватись при встановленні ціни на потенційний товар, яке передбачає аналіз ціни на товари-аналоги або товари субститути, а також аналіз рівня доходів цільової групи споживачів. Аналіз проводився експертним методом і його результати зображено в таблиці 4.20.

Таблиця 4.20 – Визначення меж встановлення цін

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
30000-50000 $/рік	50000-60000 $/рік	300000-500000 $/рік	Нижня межа – 40000 $/рік, верхня межа - 50000 $/рік

Надалі визначимо оптимальну систему збуту, в межах якого приймається рішення. Дану систему зображено в таблиці 4.21.

Таблиця 4.21 – Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Клієнт виплачує гроші на рік, при реєстрації свого пристрою в хмарному сервісі.	Забезпечити реєстрацію на оплату через онлайн сервіс	Немає посередників	Канал збуту одного рівня.

Тепер розробимо концепцію маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів. Дану концепцію зображено в таблиці 4.22.

Таблиця 4.22 – Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Клієнт намагається знайти нові методи аутентифікації.	Мережа Інтернет, соціальні мережі, відео-портали, фінансові сервіси	Інноваційність ПЗ, відносно невелика вартість ПЗ.	Продемонструвати інноваційність, простоту використання та відносну невелику вартість ПЗ.	Показати можливість не за велику ціну зацікавити користувачів.

Висновки до розділу

В даному розділі було повністю виконано перший етап розроблення стартап-проекту, а саме, виконано маркетинговий аналіз стартап-проекту.

За допомогою нього можна сказати, що існує можливість ринкової комерціалізації проекту, адже на ринку програм наявний попит на неявні системи аутентифікації, які не потребують дій користувача, та можуть попередити дії зловмисників, до того ж рентабельність роботи є досить високою.

З огляду на потенційну групу клієнтів, а саме,виробників смарт-пристроїв, та інноваційність технології є великі перспективи впровадження даного програмного забезпечення.

Для ринкової реалізації проекту доцільно обрати таку альтернативу впровадження: створення сервісу на хмарному сервері та низькорівневого інтерфейсу для фітнес пристроїв.

ВИСНОВКИ

Було проведено порівняльний аналіз різноманітних автокодувальників для біометричної верифікації користувача на основі патернів руху. Метою було визначити межі позитивного класу, щоб відрізнити конкретного користувача від решти. Неповний LSTM автокодувальник, LSTM варіаційний автокодувальник і контрактивний LSTM атокодувальник порівнювали з однокласним SVM та ізоляційним лісом. Рекурентні атокодувальник показують надійні та високі результати точності. Перевагами атокодувальників є навчання без вчителі та автоматичне генерування признаків, їх інженерія та відбір, недоліком є їх вимоги до обчислювальних ресурсів. В майбутньому ми можемо розширити перелік моделей автокодувальників та використовувати більш складні каскади автокодувальників, як, наприклад, стекінг автокодувальників різних типів в одну велику модель, проте може виникнути проблема з обчислювальною ефективністю. Тому також слід проводити дослідження щодо оптимізації з використанням моделей глибокого навчання в різних системах підтримки прийняття рішень.

Було побудовано систему біометричної верифікації користувача на базі методів машинного навчання та порівняно різні алгоритми та підходи, де найкращі результати показали штучні нейронні мережі.

ПЕРЕЛІК ПОСИЛАНЬ

1. A survey of machine learning techniques for behavioral-based biometric user authentication / Mahadi N. A. et al. Recent Advances in Cryptography and Network Security. IntechOpen. 2018. DOI: 10.5772/intechopen.76685.
2. Centeno M. P., Moorsel A. v., Castruccio S. Smartphone Continuous Authentication Using Deep Learning Autoencoders. 2017 15th Annual Conference on Privacy, Security and Trust (PST). Calgary. AB. 2017. P. 147-1478. DOI: 10.1109/PST.2017.00026.
3. Mobile Sensor Data Anonymization. Malekzadeh, Mohammad et al. Proceedings of the International Conference on Internet of Things Design and Implementation - IoTDI ’19. 2019. URL: https://arxiv.org/pdf/1810.11546.pdf (дата звернення 20.10.2018).
4. Learning Human Identity from Motion Patterns / Neverova N. et al. URL: https://arxiv.org/pdf/1511.03908.pdf (дата звернення 20.10.2018).
5. Afonso E., Aidos H., Fred A. ECG-based Biometrics using a Deep Autoencoder for Feature Learning - An Empirical Study on Transferability, 2017. DOI:10.5220/0006195404630470.
6. Shi E., Niu Y., Jakobsson M. Implicit Authentication through Learning User Behavior / ed. Burmester M., Tsudik G., Magliveras S., Ilić I. (eds) Information Security. ISC 2010. Lecture Notes in Computer Science, vol 6531. Springer, Berlin, Heidelberg, 2011. DOI:10.1007/978-3-642-18178-8_9 .
7. Continuous Authentication on Smartphones Using An Artificial Immune System / Aljohani N. et al. MAICS, 2017. URL: http://ceur-ws.org/Vol-1964/S2.pdf (дата звернення 20.10.2018).

8. Outlier detection with several methods. URL: http://scikit-learn.org/stable/auto_examples/covariance/plot_outlier_detection.html#sphx-glr-auto-examples-covariance-plot-outlier-detection-py (дата звернення 20.10.2018).
9. Поиск аномалий (Anomaly Detection). URL: https://alexanderdyakonov.wordpress.com/2017/04/19/поиск-аномалий-anomaly-detection/ (Дата звернення : 06.06.2018).
10. sklearn.svm.OneClassSVM. URL: http://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html (дата звернення 20.10.2018).
11. sklearn.ensemble.IsolationForest. URL: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html (дата звернення 20.10.2018).
12. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. URL: https://www.deeplearningbook.org/contents/autoencoders.html (дата звернення 20.10.2018).
13. Jinwon A., Sungzoon C. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. 2015. URL: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf (дата звернення 20.10.2018).
14. Fabius O., Amersfoort J. Variational Recurrent Auto-Encoders. 2014. URL: https://arxiv.org/pdf/1412.6581.pdf (дата звернення 20.10.2018).
15. UCI machine learning repository . URL: https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer (дата звернення 20.10.2018).¶

16. Casale P., Pujol O., Radeva P. Personalization and user verification in wearable systems using biometric walking patterns. Personal and Ubiquitous Computing, 16(5), 2012. P. 563-580. URL: https://www.researchgate.net/publication/227192676_Personalization_and_user_verification_in_wearable_systems_using_biometric_walking_patterns (дата звернення 20.10.2018).
17. Chollet F. Building Autoencoders in Keras. 2016. URL: https://blog.keras.io/building-autoencoders-in-keras.html (дата звернення 20.10.2018).
18. Thingom B., Rajsekhar K., Narsimhadhan A. V. Person Recognition using Smartphones' Accelerometer Data. 2017. URL: https://arxiv.org/pdf/1711.04689.pdf (дата звернення 20.10.2018).
19. GEE: A Gradient-based Explainable Variational Autoencoder for Network Anomaly Detection.Nguyen Q. et al. 2019. URL: https://arxiv.org/pdf/1903.06661.pdf (дата звернення 20.10.2018).
20. Biometric authentication technique using smartphone sensor. Laghari A. et al. 2016. DOI: 10.1109/IBCAST.2016.7429906.